Submitted to the AMIA Annual Symposium 2012.
draft material -- please do not cite

Quality Assurance in LOINC using Description Logic

Tomasz Adamusiak MD PhD[1], Olivier Bodenreider MD PhD[1]
[1] **National Library of Medicine, 8600 Rockville Pike, Bethesda, MD 20894**
olivier@nlm.nih.gov

Abstract

OBJECTIVE: To assess whether errors can be found in LOINC by changing its representation to OWL DL and comparing its classification to that of SNOMED CT. METHODS: We created Description Logic definitions for LOINC concepts in OWL and merged the ontology with SNOMED CT to enrich the relatively flat hierarchy of LOINC parts. LOINC - SNOMED CT mappings were acquired through UMLS. The resulting ontology was classified with the ConDOR reasoner. RESULTS: Transformation into DL helped to identify 427 sets of logically equivalent LOINC codes, 676 sets of logically equivalent LOINC parts, and 239 inconsistencies in LOINC multiaxial hierarchy. Automatic classification of LOINC and SNOMED CT combined increased the connectivity within LOINC hierarchy and increased its coverage by an additional 9,006 LOINC codes. CONCLUSIONS: LOINC is a well-maintained terminology. While only a relatively small number of logical inconsistencies were found, we identified a number of areas where LOINC could benefit from the application of Description Logic.

Introduction

There has been major progress both in Description Logic and ontology design since LOINC was originally developed in 1994[1]. The emergence of the standard Web Ontology Language OWL, combined with the increase in computing power removed many of the limitations that hindered early application of DL in large clinical terminologies. Terminologies developed more recently have taken adventage of DL, for example the NCI Thesaurus[2] and SNOMED CT[3].

Comprehensive clinical terminologies such as SNOMED CT tend to overlap with specialised terminologies such as LOINC, and terminological systems such as the Unified Medical Language System (UMLS) can be used to bridge between them. A relatively flat hierarchy of LOINC terms can thus be augmented with the richness of SNOMED CT relations, which can provide novel insights into the original resource[4].

Auditing clinical terminologies is an important step in assuring that they are fit for their purpose[5]. The objective is not to find errors, as any sufficiently large corpus is bound to have, but rather to identify areas for improvement. Graph-based approaches were proposed as one way to achieve this[6;7], and more recently Description Logic was suggested as a method for quality assurance in the context of SNOMED CT[8;9].

Background

LOINC is a universal standard for identifying laboratory observations. It can be considered the *lingua franca* of clinical observation exchange as it has more than 15,000 users in 145 countries[10]. It is recommended as part of the Meaningful Use and endorsed by American Clinical Laboratory Association and College of American Pathologists. A fully specified test result or clinical observation can be described formally with the following syntax: <Analyte/component>:<kind of property of observation or measurement>:<time aspect>:<system (sample)>:<scale>:-<method>[11].

Description Logic and **OWL** are a family of knowledge representation languages and OWL is specifically aimed at authoring ontologies[12]. It is endorsed by the World Wide Web Consortium (W3C) and latest specification of the language (OWL2) was released in 2009. OWL is increasingly gaining traction as a standard for implementing clinical ontologies[13].

The validation of an ontology by a DL-based classifier serves to ensure compliance with certain rules of classification, e.g., absence of cycles[3]. In general, it can be expected that the reasoner will identify two types of errors: duplicates and missing hierarchical relations[14].

The integration of LOINC and SNOMED CT through DL has been explored since 1998 [15;16]. However, prior work focused primarily on achieving the most accurate mappings between the two terminologies, which is inherently difficult due to different scopes [4]. The specific contribution of this work is the application of DL specifically to quality assurance of LOINC and to our knowledge this is the first work describing LOINC audit of any kind.

Methods

Converting LOINC into OWL

LOINC 2.36 files were created by the Regenstrief Institute. This included two additional files that contain parts names and part links to full LOINC codes that are not part of a standard LOINC distribution. We expected they would contribute significantly to our OWL version, because the links between parts and codes play an important role in determining part links among observables.

Each LOINC code was defined using relations provided by the part links file that would form a logical conjunction defining the concept. LOINC part types were translated into respective OWL object properties (see Table 1). We also created one additional relation (data property) *has_index* to model the fourth subpart of LOINC component axis. Example of a fully defined LOINC code in DL in a human-readable OWL Manchester syntax is provided in Figure 1.

LOINC_PARTS.PART_TYPE	OWL Object Property (relation)
CHALLENGE	loinc:has_challenge
CLASS	loinc:has_class
COMPONENT	loinc:has_component
DIVISORS	loinc:has_divisor
FRAGMENTS FOR SYNONYMS	loinc:has_fragments_for_synonyms
METHOD	loinc:has_method
PROPERTY	loinc:has_property
SCALE	loinc:has_scale
SUFFIX	loinc:has_suffix
SUPER SYSTEM	loinc:has_supersystem
SYSTEM	loinc:has_system
TIME	loinc:has_time_aspect
TIME MODIFIER	loinc:has_time_modifier
QUOTIENTS	loinc:has_quotients
MULTIPART	loinc:has_multipart
MULTI-AXIAL	loinc:has_multiaxial

Table 1: Breakdown of relations used in the ontology by the corresponding LOINC part type.

```
Class: 15076-3:Glucose:SCnc:Pt:Urine:Qn:

    Annotations:
        label "15076-3:Glucose:SCnc:Pt:Urine:Qn:"^^string

    EquivalentTo:
       (has_component some Glucose)
        and (has_property some SCnc)
        and (has_scale some Qn)
        and (has_system some Urine)
        and (has_time_aspect some Pt)

    SubClassOf:
       (has_class some CHEM)
        and (has_multiaxial some 'Glucose | urine'),
        'LOINC Code'
```

Figure 1: DL definition of LOINC code 15076-3 in OWL Manchester syntax.

LOINC has a multiaxial hierarchy that integrates LOINC codes and parts into a single hierarchy (see Figure 5).

Because codes and parts are different in nature, it makes sense from an ontological perspective to separate them into two different hierarchies: one for LOINC codes and abstract observations grouping them, and a separate one for LOINC parts. In practice, we want to avoid situations where a substance, e.g., glucose subsumes the observation in which this substance is also an analyte.

Pre-coordinating LOINC MULTIAXIAL parts

LOINC multiaxial hierarchy includes both parts and codes in the same graph. Most of the parts in this hierarchy are multiaxial, and can be defined in the context of their primitives, e.g., *LP:43854-6:Glucose | Urine* is a combination of COMPONENT *Glucose* and SYSTEM *Urine* parts. However, the composition of the multiaxial parts is not provided in LOINC explicitly.

We lexically matched the multiaxial parts to their primitive counterparts and created DL definitions for every part in the multiaxial hierarchy. An example is provided in Figure 2. When a lexical match was ambiguous and pointed to multiple LOINC parts, it was disambiguated based on the parts defining the underlying LOINC codes. 39,256 individual lexical matches were made in total. In 113 cases it was impossible to disambiguate the composing parts and multiple matches were accepted as valid.

We found that the string *Bld-Ser-Plas* occurring in 7,403 multiaxial part labels had no equivalent in primitive parts. It was modelled as a logical disjunction of three system parts, i.e., *has_system some LP7057-5:Blood or has_system some LP7567-3:Serum or has_system some LP7479-1:Plasma*.

```
Class: 'OBS Glucose | urine'

    Annotations:
        label "OBS Glucose | urine"^^string

    EquivalentTo:
        (has_component some Glucose)
         and (has_system some Urine)

    SubClassOf:
        'LOINC Code'
```

Figure 2: DL definition of multiaxial *LP43854-6:Glucose | Urine* LOINC part in OWL Manchester syntax.

Converting SNOMED CT into OWL

We followed the process described in the SNOMED CT Technical Implementation Guide[18]. SNOMED CT (July 2011 version) was converted into OWL by running the standard Perl transform script bundled with the distribution.

Acquiring LOINC to SNOMED CT mappings

LOINC parts were mapped to SNOMED CT terms using *owl:equivalentClass* statements. The n-to-n mappings were derived from UMLS 2011 AB release by parsing the *Concept Names and Sources File*[19]. In practice, we considered a LOINC concept equivalent to a SNOMED CT concept if both concepts were asserted under the same concept identifier in the UMLS. This provided 7,377 LOINC parts mappings to 8,161 SNOMED CT concepts. For example, *LP16699-8:Erythrocyte* and *SCT_41898006:Erythrocyte* share the same UMLS identifier *C0014792* and an equivalence axiom was asserted in the ontology accordingly.

Additionally, equivalent relations (OWL object properties) in the two terminologies were mapped with *owl:equivalentProperty* statements. It was challenging because *has_system* relation in LOINC does not have an equivalent direct relation in SNOMED CT, but rather it can be represented by a combination of relationships; linking the laboratory test first to a specimen (*has specimen*), and then linking the specimen to a substance (*specimen substance*)[4]. This

became possible to model in DL with the introduction of *owl:propertyChainAxiom* construct in OWL2. Thus, relation *hasSystem* in LOINC was asserted as equivalent to a property chain of relations *has specimen o specimen substance* in SNOMED CT.

Merging of the ontologies and classifying

The hierarchy and all the logical restrictions of SNOMED CT were preserved by merging LOINC and SNOMED CT into a single OWL ontology. The ontology was then classified using ConDOR reasoner, which was chosen because of a dramatic improvement in speed over existing ontology reasoners[20].

Computing environment

Code for parsing and serialising LOINC into OWL was written in Java7. It depends on the Java CSV library (http://opencsv.sourceforge.net) and OWL API[21] and is available under GNU lesser GPL license from: https://code.google.com/p/loinc-sem-web

Computations were performed on a dedicated server running Red Hat Enterprise Linux Server release 5.8 (Tikanga) with eight processors (Intel® Xeon™ CPU 3.20GHz) and 32GB of memory.

Results

1 LOINC and SNOMED CT overview

The final ontology consisted of 468,572 concepts in total and had over a million asserted axioms (1,577,861). 413,050 additional axioms were inferred by the reasoner (see Table 2).

	LOINC	LOINC+SNOMED CT
Number of classes	173,091	468,572
Number of asserted axioms:	677,023	1,577,861
Number of inferred axioms	126,020	413,050
LOINC codes	65,003	
LOINC parts	82,102	
	LOINC multiaxial hierarchy	
LOINC codes	47,405	
LOINC parts	25,982	

Table 2: An overview of concepts and axioms in LOINC with and without SNOMED CT.

Inferred equivalent LOINC parts

The reasoner identified 676 equivalent sets of LOINC parts comprising 1,549 LOINC parts. For example, *LP7536-8:RBC*, *LP14304-7:Erythrocytes*, and *LP16699-8:Erythrocyte* were classified by the reasoner to be equivalent to one another because they were asserted to be equivalent to the same SNOMED CT concept *SCT_41898006:Erythrocyte* via UMLS. Determining equivalent parts is an important step in the classification process for subsequent identification of equivalent LOINC concepts.

Inferred equivalent LOINC codes (intrinsic)

There were 260 sets of LOINC codes with the same definition provided by the part links table. Figure 3 provides a schematic interpretation of how two codes sharing the same part links end up with the same DL definition and are

deemed equivalent by the reasoner. These are not actual duplicates, but rather their are mistakenly linked to the same parts, for example:

```
56897-2:Cells.CD3-CD56+/100 cells:NFr:Pt:CSF:Qn
51279-8:Cells.CD3+CD56+/100 cells:NFr:Pt:CSF:Qn
```

are both linked to *LP19037-8:Cells.CD3+CD56+* and *LP35646-6:Cells.CD3-CD56+*. As is also the case with:

```
10132-9:T' wave amplitude.lead AVR:Elpot:Pt:Heart:Qn:EKG
10144-4:T wave amplitude.lead AVR:Elpot:Pt:Heart:Qn:EKG
```

both sharing the *LP31227-9:T wave amplitude.lead AVR* and *LP31243-6:T' wave amplitude.lead AVR* component parts. This is characteristic of a number of LOINC codes in the EKG.MEAS class.

Some concepts were found to be simply missing a distinguishing part link, for example:

```
64071-4:Progress note:Find:Pt:Hospital:Doc:Medical student.critical care
64072-2:Consultation note:Find:Pt:Hospital:Doc:Medical student.critical care
```

Finally, in a smaller subset no distinguishing feature could be identified and they may require additional curation:

```
46062-6:Treatments:-:Pt:^Patient:Set:
46064-2:Therapies:-:Pt:^Patient:Set:

36748-2:Views oblique:Find:Pt:Spine.cervical:Nar:XR
42164-4:Views & oblique:Find:Pt:Spine.cervical:Nar:XR

45424-9:Epilepsy:Find:Pt:^Patient:Ord:MDS
45662-4:Seizure disorder:Find:Pt:^Patient:Ord:MDS
```

This approach can be considered intrinsic to LOINC as the aforementioned results could be achieved without Description Logic by simply running a detailed database query against the LOINC codes table.

Inferred equivalent LOINC codes (extrinsic)

However, once LOINC is augmented with additional information from SNOMED CT (see *Acquiring LOINC to SNOMED CT mappings*) it becomes possible to validate LOINC assertions externally. This extrinsic approach allowed to identify additional 167 sets of equivalent LOINC codes. Figure 4 demonstrates in more detail how this approach differs from the aforementioned intrinsic inference. Examples in this category include:

```
10374-7:Helmet cells:ACnc:Pt:Bld:Ord:Microscopy.light
800-3:Schistocytes:ACnc:Pt:Bld:Ord:Microscopy.light

8703-1:Physical findings:Find:Pt:Extremities:Nom:Observed
32430-1:Physical findings:Find:Pt:Extremity:Nom:Observed

39037-7:Multisection^W contrast IV:Find:Pt:Upper extremity:Nar:MRI
36208-7:Multisection^W contrast IV:Find:Pt:Upper arm:Nar:MRI
```

This method identified codes differing only in grammatical number (*Extremities* vs. *Extremity*), codes with synonymous components (*Helmet cells* vs. *Schistocytes*), as well as parts that could be considered synonymous (*Upper extremity* vs. *Upper arm*) depending on the context. However, it did identify some codes incorrectly as equivalent:

```
9105-8:Fluid intake.total:VRat:8H:^Patient:Qn:
9259-3:Fluid output.total:VRat:8H:^Patient:Qn:
```

Complete data sets are available as Supplementary Information at http://goo.gl/3Opk2 (intrinsic) and http://goo.gl/ZI10T (extrinsic).

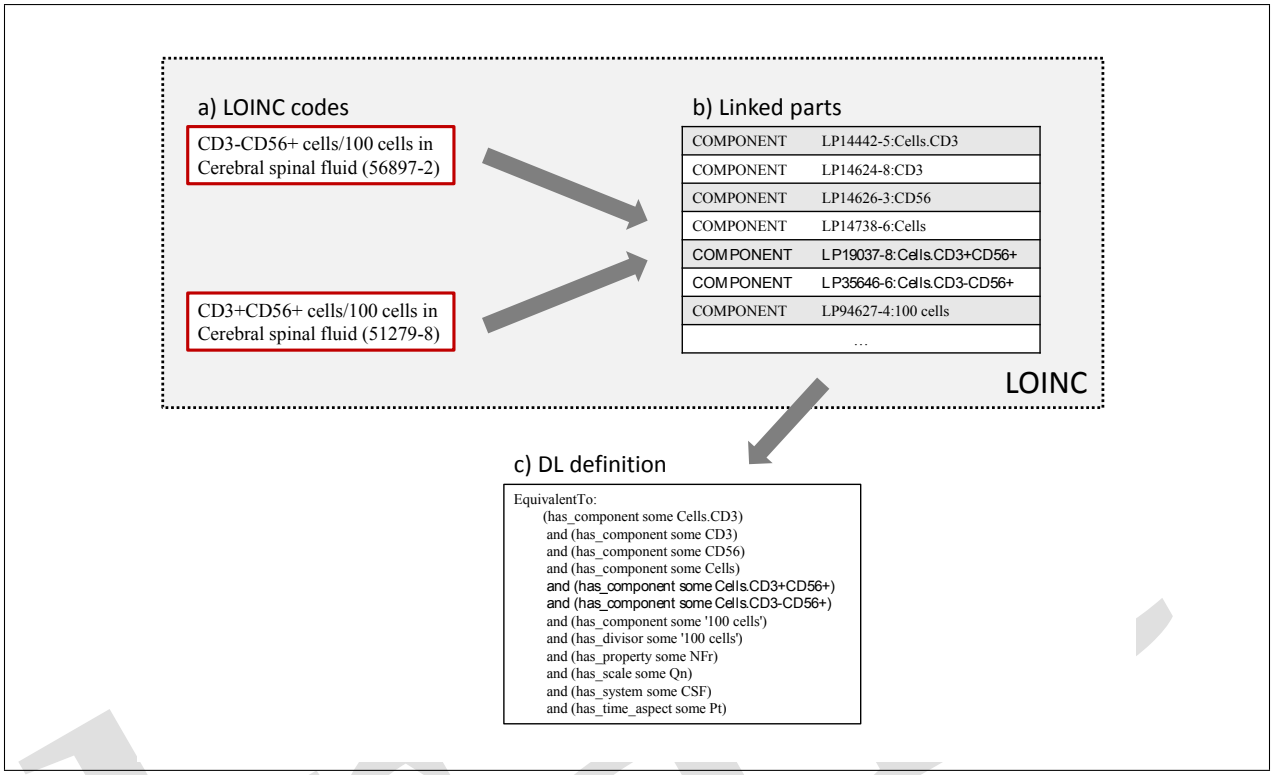


Figure 3: LOINC codes inferred as equivalent due to shared links (intrinsic information).

Inferred multiaxial hierarchy

There are 73,387 concepts in the multiaxial hierarchy, which covers 47,461 (73%) LOINC codes terms and 26,834 (32%) LOINC parts. 17,598 LOINC codes are outside of the hierarchy (see Table 2).

The reasoner starting from a pre-coordinated list of abstract observations and LOINC codes created a hierarchy with 56,411 LOINC codes and essentially enhanced the original hierarchy by 9,006 additional codes. General characteristics of the two graphs: original multiaxial and inferred are presented in Table 3 and a neighbourhood of glucose tests in urine is shown in Figures 5 and 6 for the multiaxial and inferred hierarchies respectively.

	LOINC	**Inferred**
Number of nodes	73,387	82,350
Network diameter	15	13
Connected components	8	513
Shortest paths	425,976	1,119,232
Characteristic path length	3.39	3.81
Average number of neighbours	2.01	3.40

Table 3: Network analysis comparing LOINC multiaxial and inferred hierarchies.

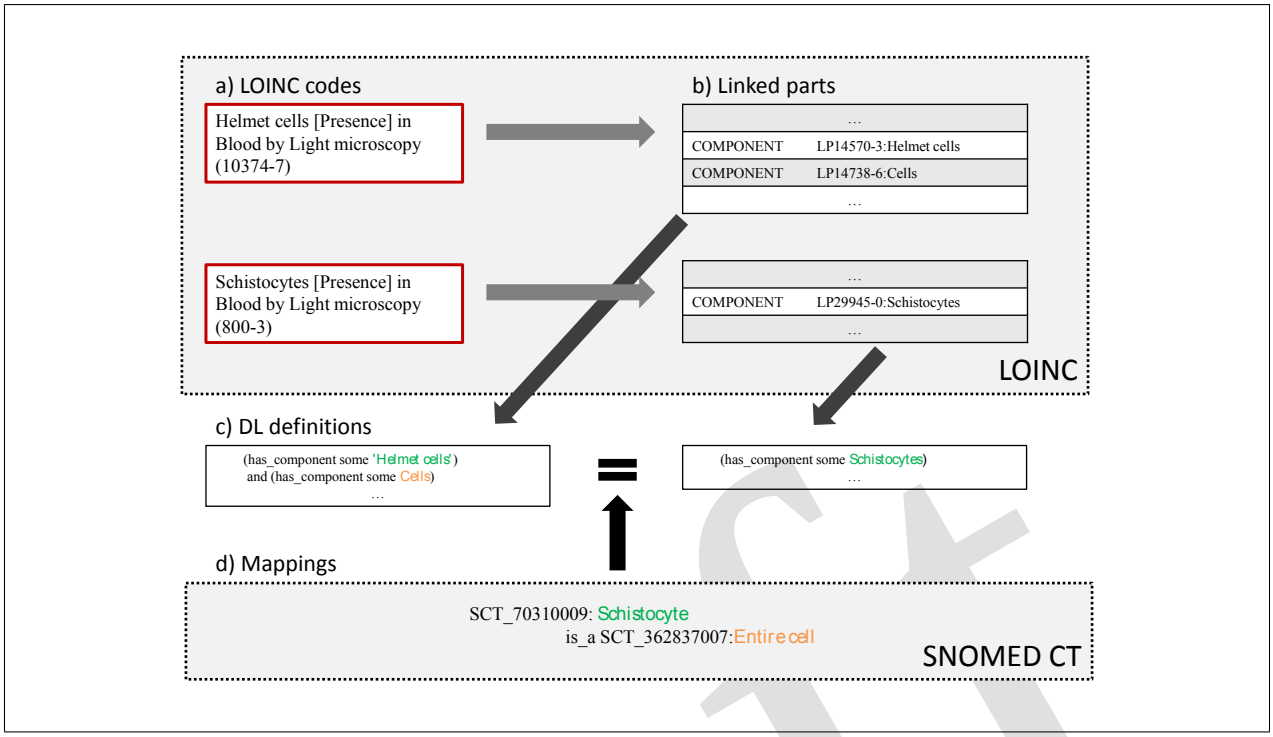


Figure 4: LOINC codes inferred as equivalent via SNOMED CT (extrinsic information).

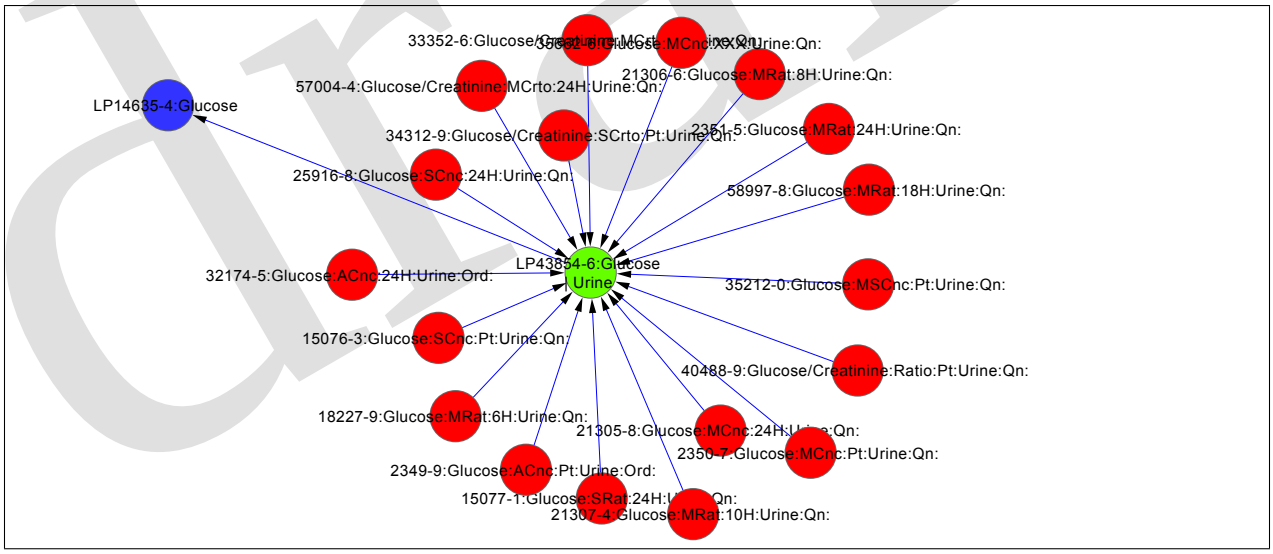


Figure 5: Neighbourhood of glucose tests in urine in the multiaxial hierarchy. LOINC codes are shown in red, MULTIAXIAL *LP43854-6:Glucose| Urine* in green, and COMPONENT *LP14635-4:Glucose* in blue. Edges represent *is_a* (*subClassOf*) relations. [created in Cytoscape [17]]

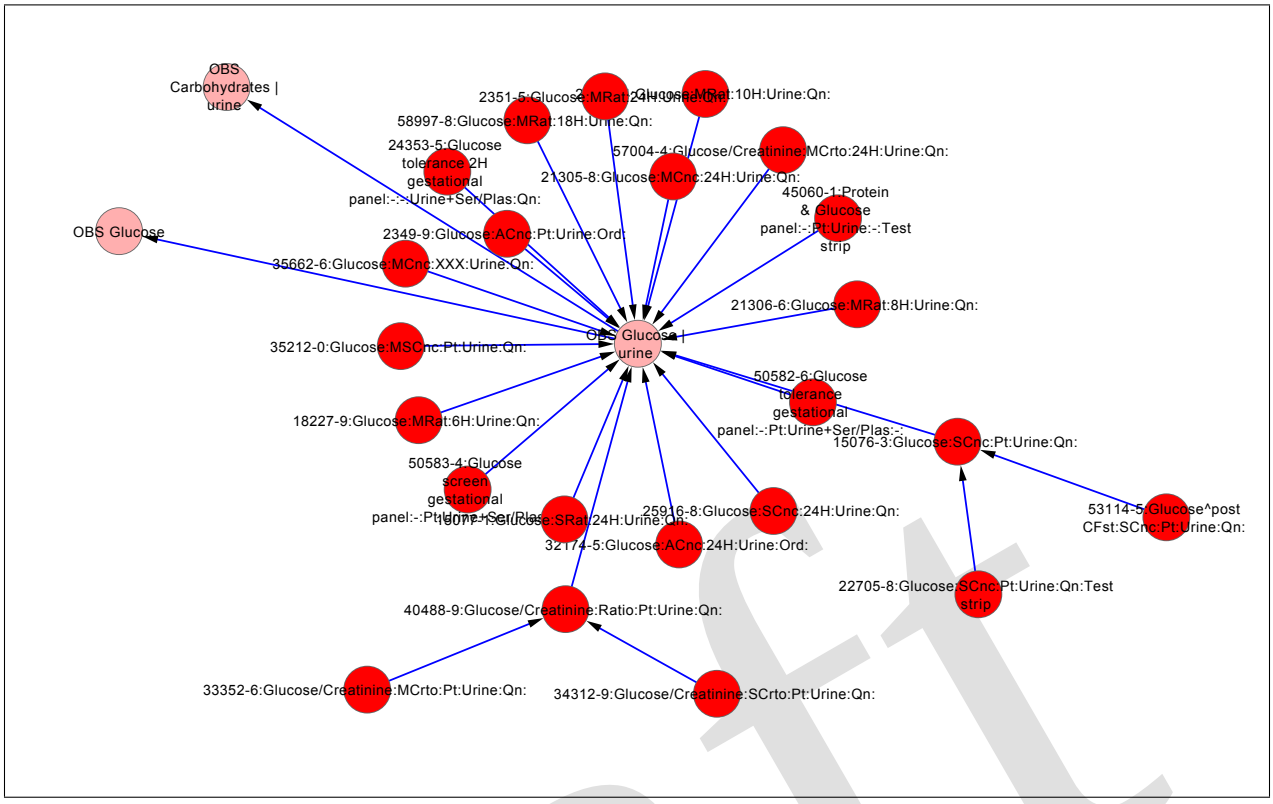
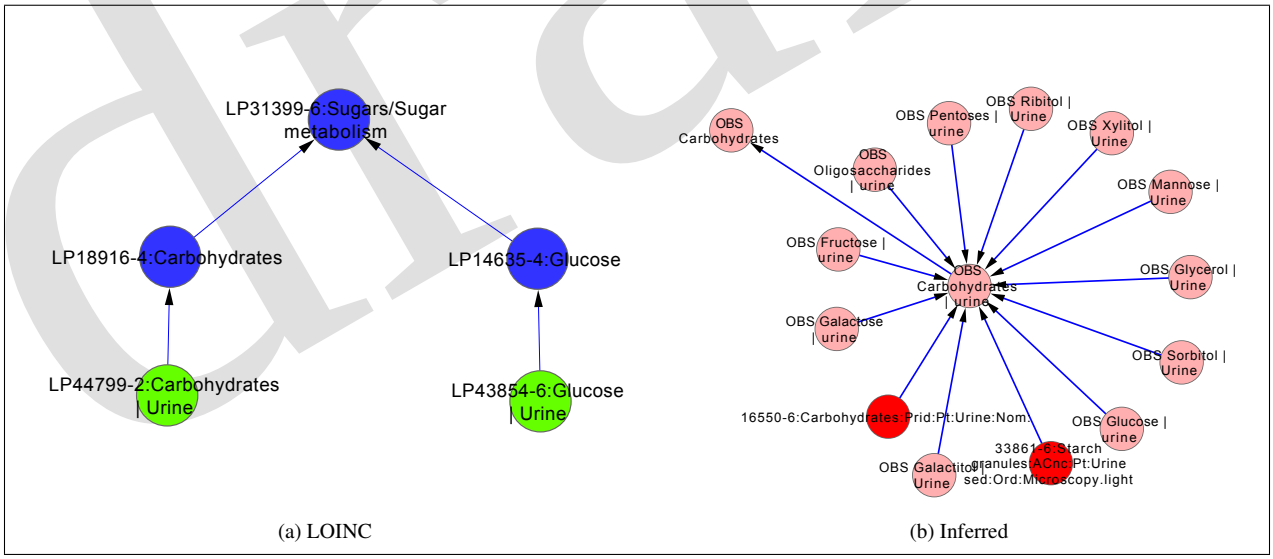


Figure 6: Neighbourhood of glucose tests in urine in the inferred hierarchy (compare with Figure 5. LOINC codes are shown in red and grouping observations, e.g., *OBS Glucose | Urine* in pink. Edges represent *is_a* (*subClassOf*) relations. [created in Cytoscape [17]]



(a) LOINC (b) Inferred

Figure 7: Paths between *Carbohydrates | Urine* and *Glucose | Urine* observations in LOINC multiaxial and inferred hierarchies. LOINC codes are shown in red, abstract observations in pink, MULTIAXIAL parts in green, and COMPONENT parts in blue. Edges represent *is_a* (*subClassOf*) relations. [created in Cytoscape [17]]

Figure 6 demonstrates two important aspects of the inferred hierarchy. There is a new access point to *Glucose | Urine* codes via *Carbohydrates | Urine*. Figure 7a shows that this path does not exist in the original LOINC hierarchy, where the two abstract observations are sharing the same parent *Sugar metabolism*, but otherwise are not in direct relationship to one another. Furthermore, some codes are now subsumed as more specific, e.g., *22705-8:Glucose:SCnc:Pt:Urine:Qn:Test strip* became a child of a more general *15076-3:Glucose:SCnC:pt:Urine:Qn:* test.

Inconsistencies in LOINC multiaxial hierarchy

The reasoner identified 198 sets of equivalent multiaxial parts. No concepts already asserted in the multiaxial hierarchy were excluded by the reasoner. 122 discouraged LOINC codes already in the hierarchy were omitted from the analysis. 239 LOINC codes were found to be incorrectly asserted in respect to their parts pre-coordination. In majority of cases the assertions were correct lexically, but the codes lacked a link to more specific part. These codes predominantly (183 concepts) were of scale type Document, e.g., *28626-0:History and physical note:Find:Pt:Setting:Doc:Physician* was asserted under *History and physical note*, but the reasoner placed it under a more general observation *Note* (Document Ontology branch), because *28626-0* was defined as *has_component some Note*. Another class of findings were chemical compounds, e.g., *38639-1:Boron trifluoride:MCnc:Pt:Air:Qn:* was asserted under *Boron trifluoride | air*, but the reasoner inferred it under *Boron* as it was missing the more specific *Boron trifluoride* connection.

Only one class of true errors was identified. A number of codes for measurements of fatty acids, e.g., *44084-2:Fatty acids:Imp:Pt:Ser/Plas:Nom:* were originally asserted under *7-hydroxyoctanoate | Urine* (note that both component and system assertions were wrong), and the reasoner correctly inferred them elsewhere, in this particular case under *Lipids | bld-ser-plas*.

A full list is available as Supplementary Information at http://goo.gl/E97O4.

Discussion

Significance

1. Error detection
   (a) Duplicates
   We found that while the DL approach does in fact identify potential duplicates it is much more sensitive to insufficiencies in modelling.
   (b) Missing hierarchical relations
   The reasoner was successful in identifying a significant number of missing connections between the LOINC codes. A comparison of the two networks (see Table 3) confirms that the inferred hierarchy is indeed better connected as it has more paths and neighbours.
   (c) Inconsistencies in hierarchy
   When LOINC codes were identified to be incorrectly asserted in the hierarchy, it was most likely due to insufficient modelling rather than erroneous assertions. However, some errors might have been more difficult to detect due to inconsistencies in parts hierarchy. For example, *LP51132-6:N-methyl valine | Bld-Ser-Plas* is asserted directly under *LP501109-5:Amino Acids | Amniotic fluid*. When the reasoner takes this assertion at face value, it fails to identify the clash in systems (*Bld-ser-Plas* vs. *Amniotic fluid*) between the two parts.
2. Enhanced navigation
   In the current multiaxial hierarchy, LOINC codes have at most one parent, though the abstract observations are interlinked more. The DL representation provides access to more LOINC codes and more paths between the codes in the hierarchy, which directly translates into more ways a particular code can be discovered. By adding new paths to the hierarchy it also enables queries that were otherwise not possible. Figure 7b demonstrates this on the example of *Carbohydrates | Urine*, which in the inferred hierarchy not only returns a set of *Glucose | Urine* tests, but also all other carbohydrate measurements in urine.

3. Enhanced subsumption

   The new paths are not limited to abstract observations that group LOINC codes, but also LOINC codes themselves can be in direct relationships. This actually limits the need for creating abstract observations to group them.

4. Maintenance

   A typical scenario where a new LOINC code is requested by an external user requires several manual and error-prone steps to add the code. Firstly, you need to confirm that the test or its close variant does not exist already in the terminology. This requires several queries and expert knowledge on the actual naming conventions in a particular field. Secondly, you need to identify the best place in the hierarchy to place the new term.

   DL approach can to a large extent simplify both steps through automated classification. There is also now standard tooling to work and manipulate OWL ontologies, such as OWL API and Protégé [22].

Intrinsic vs. extrinsic approach

It is important for the logical assertions in the ontology to be externally validated as the resulting inference is only as good as the original assertions. An external resource can also shed some lights on areas that were not modelled sufficiently.

Recommendations

1. Create logical definitions for codes

   Description Logic has long been recognised as indispensable to achieving convergence of clinical terminologies [23]. Having logical definitions for codes enables the "lexically assign, logically refine" strategy followed by other clinical terminologies such as SNOMED CT [9]. Transition to DL could be considered a natural consequence as LOINC codes already have multiaxial composition.

2. Have an inferred hierarchy

   The hierarchy of parts could be taken directly from SNOMED CT thus minimising the effort involved. A hierarchy of parts would mean that hierarchy of codes could be inferred automatically.

3. Parts vs. codes

   Multiaxial codes represent abstract groupings of observations and as such logically are no longer parts. Parts in general should not be in hierarchical relation to codes, which is especially true if they are at the same time used to define them.

4. Alignment with SNOMED CT

   Consider the examples listed in Table 4, where a single SNOMED CT concept maps to several LOINC parts. What does it mean to have several parts in LOINC map to SNOMED CT? If indeed they are different concepts that their name should not represent the same entities.

Limitations

Relying on UMLS to provide mappings between SNOMED CT and LOINC may lead to unintended logical consequences [24] and the mappings are not guaranteed to be one-to-one [25]. We manually surveyed a sample of the inferred mappings and were not able to identify any major issues. Nevertheless, imposing a specific ontological commitment onto LOINC might not necessarily be desirable and can produce unintended results, such as an equivalence between *Upper extremity* and *Upper arm*.

SCT_385421009:Site of distant metastasis	
COMPONENT	LP73362-3:Sites of distant metastasis
COMPONENT	LP73358-1:Site of distant metastasis
COMPONENT	LP72485-3:Distant metastasis site
SCT_41040004:Complete trisomy 21 syndrome	
COMPONENT	LP19800-9:Chromosome 21 trisomy
COMPONENT	LP74785-4:Down's syndrome
FRAGMENTS FOR SYNONYMS	LP70386-5:Trisomy 21
SCT_3711007:Structure of great blood vessel (organ)	
SYSTEM	LP7303-3:Heart.great vessels
SYSTEM	LP33690-6:Great vessel
SYSTEM	LP30622-2:Great vessels
SCT_66019005:Limb structure	
COMPONENT	LP121777-9:Extremity
SYSTEM	LP7216-7:Extremities
SYSTEM	LP7395-9:Limbs
SYSTEM	LP29945-0:Extremity
SCT_40733004:Infectious disease	
COMPONENT	LP74970-2:Infections
COMPONENT	LP74968-6:Infection unspecified
COMPONENT	LP128526-3:Infectious disease
COMPONENT	LP97560-4:Infectious diseases
METHOD	LP32901-8:Infectious disease
SCT_245543004:Dentition	
SYSTEM	LP115763-7:Tooth
SYSTEM	LP76036-0:Tooth
SYSTEM	LP40376-3:Teeth
SYSTEM	LP35154-1:Dentition
COMPONENT	LP74717-7:Dentition
COMPONENT	LP115762-9:Tooth
SCT_83578000:Surgical	
COMPONENT	LP73443-1:Surgery procedure
COMPONENT	LP73010-8:Operations (?)
COMPONENT	LP73440-7:Surgery
METHOD	LP32925-7:Surgery
SYSTEM	LP7616-8:Surgical procedure
FRAGMENTS FOR SYNONYMS	LP100628-9:Surgical
CHALLENGE	LP97910-1:Surgery
CLASS	LP7802-4:H&P.SURG PROC

Table 4: Examples of LOINC - SNOMED CT multiple alignments.

Modelling LOINC codes with conjunctions is suboptimal for more complex observations, where disjunctions would be more appropriate. This is however challenging to automate due to some disconnection between lexical descriptions and parts compositions for more complex terms.

Glossary

OWL	W3C Web Ontology Language
DL	Description Logic
HL7	Health Level Seven

Acknowledgements

We would like to thank Dr. Clement McDonald for helpful comments and Regenstrief Institute for providing the necessary files. This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM) and the The Oak Ridge Institute for Science and Education (ORISE) Training Program in Clinical Informatics managed for the U.S. Department of Energy (DOE) by Oak Ridge Associated Universities (ORAU).

References

[1] a W Forrey, C J McDonald, G DeMoor, S M Huff, D Leavelle, D Leland, T Fiers, L Charles, B Griffin, F Stalling, A Tullis, K Hutchins, and J Baenziger. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. *Clinical chemistry*, 42(1):81–90, January 1996. ISSN 0009-9147. URL http://www.ncbi.nlm.nih.gov/pubmed/8565239.

[2] Frank W Hartel, Sherri de Coronado, Robert Dionne, Gilberto Fragoso, and Jennifer Golbeck. Modeling a description logic vocabulary for cancer research. *Journal of biomedical informatics*, 38(2):114–29, April 2005. ISSN 1532-0464. doi: 10.1016/j.jbi.2004.09.001. URL http://www.ncbi.nlm.nih.gov/pubmed/15797001.

[3] Olivier Bodenreider, Barry Smith, Anand Kumar, and Anita Burgun. Investigating subsumption in SNOMED CT: an exploration into large description logic-based biomedical terminologies. *Artificial intelligence in medicine*, 39(3):183–95, March 2007. ISSN 0933-3657. doi: 10.1016/j.artmed.2006.12.003. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2442845&tool=pmcentrez&rendertype=abstract.

[4] Olivier Bodenreider. Issues in mapping LOINC laboratory tests to SNOMED CT. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, pages 51–5, January 2008. ISSN 1942-597X. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2655945&tool=pmcentrez&rendertype=abstract.

[5] J Geller, Y Perl, M Halper, and R Cornet. Special issue on auditing of terminologies. *Journal of biomedical informatics*, 42(3):407–11, June 2009. ISSN 1532-0480. doi: 10.1016/j.jbi.2009.04.006. URL http://www.ncbi.nlm.nih.gov/pubmed/19465342.

[6] Lee Peters and Olivier Bodenreider. Using the RxNorm web services API for quality assurance purposes. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, pages 591–5, January 2008. ISSN 1942-597X. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2656097&tool=pmcentrez&rendertype=abstract.

[7] Olivier Bodenreider and Lee B Peters. A graph-based approach to auditing RxNorm. *Journal of biomedical informatics*, 42(3):558–70, June 2009. ISSN 1532-0480. doi: 10.1016/j.jbi.2009.04.004. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2722378&tool=pmcentrez&rendertype=abstract.

[8] Alan L Rector, Sam Brandt, and Thomas Schneider. Getting the foot out of the pelvis: modeling problems affecting use of SNOMED CT hierarchies in practical applications. *Journal of the American Medical Informatics Association : JAMIA*, 18(4):432–40, 2011. ISSN 1527-974X. doi: 10.1136/amiajnl-2010-000045. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3128394&tool=pmcentrez&rendertype=abstract.

[9] Alan Rector and Luigi Iannone. Lexically suggest, logically define: Quality assurance of the use of qualifiers and expected results of post-coordination in SNOMED CT. *Journal of biomedical informatics*, October 2011. ISSN 1532-0480. doi: 10.1016/j.jbi.2011.10.002. URL http://www.ncbi.nlm.nih.gov/pubmed/22024315.

[10] C. J. McDonald. LOINC, a Universal Standard for Identifying Laboratory Observations: A 5-Year Update. *Clinical Chemistry*, 49(4):624–633, April 2003. ISSN 0009-9147. doi: 10.1373/49.4.624. URL http://www.clinchem.org/cgi/content/abstract/49/4/624.

[11] C McDonald, SM Huff, J Suico, and K. Mercer. Logical Observation Identifiers Names and Codes (LOINC R) users' guide. *Indianapolis: Regenstrief Institute*, 2004.

[12] Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003. ISBN 0521781760.

[13] Steffen Staab and Rudi Studer. *Handbook on Ontologies*. Springer, 2009. ISBN 3540709991.

[14] Duo Wei and Olivier Bodenreider. Using the abstraction network in complement to description logics for quality assurance in biomedical terminologies - a case study in SNOMED CT. *Studies in health technology and informatics*, 160(Pt 2):1070–4, January 2010. ISSN 0926-9630. URL http://www.ncbi.nlm.nih.gov/pubmed/20841848.

[15] R H Dolin, S M Huff, R A Rocha, K A Spackman, and K E Campbell. Evaluation of a "lexically assign, logically refine" strategy for semi-automated integration of overlapping terminologies. *Journal of the American Medical Informatics Association : JAMIA*, 5(2):203–13, 1998. ISSN 1067-5027. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=61291&tool=pmcentrez&rendertype=abstract.

[16] K A Spackman. Integrating sources for a clinical reference terminology: experience linking SNOMED to LOINC and drug vocabularies. *Studies in health technology and informatics*, 52 Pt 1:600–3, January 1998. ISSN 0926-9630. URL http://www.ncbi.nlm.nih.gov/pubmed/10384525.

[17] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S Baliga, Jonathan T Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–504, November 2003. ISSN 1088-9051. doi: 10.1101/gr.1239303. URL http://genome.cshlp.org/cgi/content/abstract/13/11/2498.

[18] SNOMED CT® Technical Implementation Guide January 2012 International Release (US English). URL http://ihtsdo.org/fileadmin/user_upload/doc/download/doc_TechnicalImplementationGuide_Current-en-US_INT_20110731.pdf.

[19] UMLS® Reference Manual, September 2009. URL http://www.ncbi.nlm.nih.gov/books/NBK9676/.

[20] Yevgeny Kazakov. Consequence-Driven Reasoning for Horn SHIQ Ontologies. In *Proceedings of the 21st International Conference on Artificial Intelligence (IJCAI 2009)*, pages 2040–2045, 2009. URL http://ijcai.org/papers09/Papers/IJCAI09-336.pdf.

[21] Matthew; Bechhofer Sean Horridge. The OWL API: A Java API for Working with OWL 2 Ontologies. In *OWLED 2009, 6th OWL Experienced and Directions Workshop, Chantilly, Virginia, October 2009*, 2009.

[22] Daniel L Rubin, Natalya F Noy, and Mark A Musen. Protégé: a tool for managing and using terminology in radiology applications. *Journal of digital imaging : the official journal of the Society for Computer Applications in Radiology*, 20 Suppl 1:34–46, November 2007. ISSN 0897-1889. doi: 10.1007/s10278-007-9065-0. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2039856&tool=pmcentrez&rendertype=abstract.

[23] K A Spackman and K E Campbell. Compositional concept representation using SNOMED: towards further convergence of clinical terminologies. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, pages 740–4, January 1998. ISSN 1531-605X. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2232179&tool=pmcentrez&rendertype=abstract.

[24] Ernesto Jiménez-Ruiz, Bernardo Cuenca Grau, Ian Horrocks, and Rafael Berlanga. Logic-based assessment of the compatibility of UMLS ontology sources. *Journal of biomedical semantics*, 2 Suppl 1:S2, January 2011. ISSN 2041-1480. doi: 10.1186/2041-1480-2-S1-S2. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3105494&tool=pmcentrez&rendertype=abstract.

[25] Prakash M Nadkarni and Jonathan D Darer. Determining correspondences between high-frequency MedDRA concepts and SNOMED: a case study. *BMC medical informatics and decision making*, 10(1):66, January 2010. ISSN 1472-6947. doi: 10.1186/1472-6947-10-66. URL http://www.biomedcentral.com/1472-6947/10/66.